فردحسینی، مهسا؛ طاهری، سید مهدی؛ حریری، نجلا؛ باب‌الحوائجی، فهیمه؛ نوشین‌فرد، فاطمه (۱۳۹۹). بررسی میزان همخوانی انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با فهرست .... پژوهش‌نامه کتابداری و اطلاع‌رسانی، ۱۰(۲)، ۲۲۹-۲۴۷. DOI: 10.22067/infosci.2021.24040.0

بررسی میزان همخوانی انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با فهرست اصطلاحات نوع خاص منبع استاندارد مارک ۲۱

مهسا فردحسینی[۱]، سید مهدی طاهری[۲]، نجلا حریری[۳]، فهیمه باب‌الحوائجی[۴]، فاطمه نوشین‌فرد[۵]

تاریخ دریافت: ۱۳۹۸/۱۰/۱۷ تاریخ پذیرش: ۱۳۹۹/۰۳/۲۴ نوع مقاله: پژوهشی

چکیده

مقدمه: پژوهش حاضر با هدف تبیین میزان همخوانی انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با بازنمایی آن‌ها در فهرست اصطلاحات نوع خاص منبع استاندارد مارک ۲۱ انجام شد.

روش‌شناسی: این پژوهش از نظر هدف، یک پژوهش کاربردی است که بر اساس روش تحلیل محتوا صورت گرفت و جامعه پژوهش را انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند تشکیل می‌دادند. برای گردآوری داده‌ها، از ابزار سیاهه وارسی مبتنی بر فهرست اصطلاحات نوع خاص منبع مارک، برای بررسی میزان همخوانی با انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند، و روش مشاهده طراحی شده استفاده شد که برای این منظور در ابتدا هر یک از موجودیت‌های آثار خلاقانه و سپس موجودیت‌های دیگر در فرانمای تولید داده‌های ساختارمند با اصطلاحات نوع خاص منبع در استاندارد «مارک» مورد تحلیل، و انواع موجودیت‌های داده‌ای مرتبط شناسایی و تعیین شدند.

یافته‌ها: یافته‌های این پژوهش نشان داد که ۶۴ درصد از اصطلاحات نوع خاص منبع در مارک وجود دارد که در فرانمای تولید داده‌های ساختارمند هیچ نوع موجودیت داده‌ای معادل آنها تعریف نشده، و تنها ۳۶ درصد اصطلاحات نوع خاص منبع مارک با فرانمای تولید داده‌های ساختارمند همخوان است. بدین معنی که برای ۱۰۰ اصطلاح نوع خاص منبع مارک فقط ۳۶ معادل در فرانمای تولید داده‌های ساختارمند وجود داشت. در این میان، ۲۴ اصطلاح از اصطلاحات نوع خاص منبع مارک با موجودیت

۱. دانش‌آموخته دکتری علم اطلاعات و دانش‌شناسی واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران، mahsafardehoseini@gmail.com

۲. دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشگاه علامه طباطبائی، تهران، ایران، (نویسنده مسئول) taherismster@gmail.com

۳. استاد گروه علم ارتباطات و دانش‌شناسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

۴. دانشیار گروه علم ارتباطات و دانش‌شناسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

۵. دانشیار گروه علم ارتباطات و دانش‌شناسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

آثار خلاقانه در فرانمای تولید داده‌های ساختارمند همخوانی کامل و دقیق داشتند و ۱۲ اصطلاح با سایر موجودیت‌های فرانما همخوان بودند.

نتیجه: برای توصیف و سازماندهی اشیای دیجیتالی، به دلیل آنکه بیشتر موجودیت‌های داده‌ای موجود در وب، نسخه دیجیتالی شده اشیای محتوایی فیزیکی هستند، بنابراین ویژگی‌های هر دو نسخه باید مدنظر قرار گیرد. همچنین با توجه به اینکه فرانمای تولید داده‌های ساختارمند بخشی از موجودیت‌های داده‌ای بافت کتابشناختی را در نظر نگرفته، در نتیجه این فرانما چنان چه بخواهد نگاره دانش را تکمیل کند، باید برای توصیف تمام اصطلاحات نوع خاص منبع بافت میراث فرهنگی (از جمله بافت کتابشناختی) فرانمای مناسب و ویژه طراحی نماید. نیز توصیه می‌شود، اصلاحات و واژگان تخصصی سایر بخش‌های بافت میراث فرهنگی (بافت موزه و آرشیو) با موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند تطبیق داده شوند تا زمینه برای بازنمون کامل دانش در بافت میراث فرهنگی فراهم گردد.

کلیدواژه‌ها: آثار خلاقانه، بافت کتابشناختی، بافت میراث فرهنگی، فرانمای تولید داده‌های ساختارمند، فهرست اصطلاحات نوع خاص منبع مارک

مقدمه

شبکه وب، به‌عنوان بزرگترین محیط ذخیره و بازیابی داده‌ها، دربرگیرنده انبوهی از داده‌های ساختاریافته و ساختارنیافته است. این داده‌ها برای دسترس‌پذیر شدن نیازمند توصیف و پردازش هستند. برای توصیف منابع (اشیای محتوایی)، واژگان‌های کنترل شده زیادی وجود دارد که می‌توان از آنها برای توصیف نوع خاص منبع[1] یک اثر استفاده نمود. واژگان نوع استاندارد فراداده‌ای هسته دابین[2] یک نمونه از چنین واژگان‌های کنترل شده است. این مقادیر شامل تصویر، صدا و متن است که می‌تواند برای توصیف انواع موجودیت‌های یک منبع مورد استفاده قرار گیرد. فهرست‌نویسان برای توصیف اشیای محتوایی، فهرستی بیش از ۶۰ اصطلاح کنترل شده را که برای این منظور تأیید شده در اختیار دارند. پژوهشگران فهرست دسترس‌پذیر همگانی (اوسی‌ال‌سی)[3] نیز با شناخت ارزش اصطلاحات نوع خاص منبع برای کشف اشیای محتوایی، مجموعه‌ای از خدمات وبی[4] را برای اصطلاحات نوع خاص منبع به‌عنوان بخشی از اصطلاحات اوسی‌ال‌سی نمونه‌سازی کرده‌اند (Vizine-Goetz, Childress & Houghton, 2005). در بافت میراث فرهنگی، استانداردهای فراداده‌ای، استانداردهای محتوایی و الگوهای مفهومی گوناگونی برای مدیریت اشیای محتوایی این بافت متناسب با نیازهای کاربران از جمله یافتن، شناسایی، انتخاب، دسترسی، کشف و

1. Genre
2. DCMI Type Vocabulary
3. OCLC
4. Web Services

جز آن طراحی شده‌اند تا کاربران بتوانند به بهترین روش به اشیای محتوایی دسترسی یابند. استاندارد فراداده‌ای مارک از جمله قوی‌ترین و غنی‌ترین استانداردهای فراداده‌ای موجود در بافت کتابشناختی است که تا حدودی برای توصیف اشیای دیگر بافت‌های جامعه میراث فرهنگی نیز سازگار شده است. با وجود طراحی استانداردهای فراداده‌ای، استاندارد فراداده‌ای مارک همچنان پیشتاز است، به نحوی که با پیدایش استاندارد محتوایی آردی‌اِی هنوز هم کتابخانه کنگره نظام خود را تغییر نداده، و بر اساس دستورالعمل‌ها و رهنمودهای جدید این استاندارد، فیلدهایی را به مارک می‌افزاید[1]. فهرست اصطلاحات نوع خاص منبع مارک[2] نیز برای پوشش انواع منابع اطلاعاتی در بافت کتابشناختی از کامل‌ترین فهرست‌ها برای نوع خاص منبع این بافت است. این فهرست شامل اصطلاحاتی است که دارای دسته‌بندی‌های عمومی[3] (نوع عام منبع)، و گروه‌بندی نوع خاص منبع‌هایی که می‌توانند در انواع محیط‌های اطلاعاتی به کار روند، و از ارزش‌های اصطلاحات کنترل شده عناصر داده‌ای مختلف در قالب مارک ۲۱ برای داده‌های کتابشناختی گردآوری شده است (MARC Genre Term List[4], 2020). بر مبنای چنین تجربه‌ها و تلاش‌هایی، بدیهی است که الگوها و استانداردهای کتابخانه‌ای به دلیل پیشینه غنی و تجربه گسترده در استفاده از فراداده‌ها، به بلوغ نسبی در توصیف موجودیت‌های داده‌ای در محیط شبکه‌ای و غیرشبکه‌ای رسیده‌اند. از سوی دیگر، طراحان موتورهای کاوش با هدف ساختارمند کردن داده‌ها برای انجام جستجوهای معنایی و دسترسی بهتر و دقیق‌تر به اشیای محتوایی وب، فعالیت‌هایی را در تولید واژگان برای توصیف اشیای محتوایی انجام داده‌اند. یکی از جدیدترین این کوشش‌ها، تدوین فرانمای تولید داده‌های ساختارمند[5] توسط طراحان موتورهای کاوش گوگل، یاهو، بینگ، و یاندکس است که بر پایه رویکرد فراداده‌ای و نزدیک به ویژگی‌های بافت فراداده‌ای میراث فرهنگی، نیز در راستای توسعه وب معنایی انجام شده است. طراحان موتور کاوش گوگل اذعان می‌دارند که فرانمای تولید داده‌های ساختارمند، داده‌هایی با ساختارهای معناشناختی سطح بالا هستند، و این ساختارمندی

---

۱. اگرچه کتابخانه کنگره در حال توسعه چهارچوب کتابشناختی (بیب‌فریم) مبتنی بر روش داده‌های پیوندی به‌عنوان جایگزین مارک است، اما با وجود طراحی نسخه دوم این چهارچوب، به دلایل گوناگون، همچنان قالب مارک مبنای توصیف و مدیریت داده‌های کتابخانه‌ای است.

۲. فهرست اصطلاحات نوع خاص منبع مارک به‌عنوان یکی از واژگان‌های خدمات داده‌های پیوندی مارک، در نشانی http://id.loc.gov/vocabulary/marcgt.html ارائه شده، و از سوی استاندارد چهارچوب کتابشناختی (بیب‌فریم) و استاندارد توصیف دسترسی به منبع (آردی‌اِی) نیز برای صفت GenreForm توصیه گردیده است.

3. General Categories

4. https://www.loc.gov/standards/valuelist/marcgt.html

5. schema.org

موتورهای کاوش را قادر می‌سازد، داده‌های هدف را به‌صورت معنایی پردازش (نمایه‌سازی) نموده، و نمایش دهند (Google developers, 2017). فرانمای تولید داده‌های ساختارمند یک فعالیت مشارکتی_اجتماعی برای ایجاد، نگهداری و ارتقای فرانماها برای تولید داده‌های ساختارمند در وب، پیام‌های پست الکترونیکی و غیره است، و واژگان‌های[1] آن در بسترهای نحوی نشانه‌گذاری معنایی گوناگونی شامل قالب‌های خرد[2]، داده‌های خرد[3]، چهارچوب توصیف منبع در ویژگی‌ها (آردی‌اف.ای)[4]، و نشانه‌گذاری اشیاء مبتنی بر جاوا اسکریپت (جیسون_ال‌دی)[5] استفاده می‌شود. این واژگان‌ها، موجودیت‌ها، روابط بین آن‌ها و کارکردهای پیشینه‌های فراداده‌ای را پوشش می‌دهند و به راحتی قابل گسترش‌اند. در فرانمای تولید داده‌های ساختارمند، دو ساختار سلسله مراتب تعریف شده است: یکی برای ارزش‌های متنی صفت‌ها، و دیگری برای اشیایی (موجودیت‌های داده‌ای)[6] که توسط این ارزش‌ها توصیف می‌شوند (Schema.org, 2019).

همچنین به منظور توسعه و بهبود کارکردهای مورد انتظار کاربران از نظام‌های بازیابی اطلاعات، طراحان موتور کاوش گوگل، طراحی و اجرای پروژه‌ای را با نام نگاره دانش[7] آغاز کرده‌اند که همانند الگوهای مفهومی بافت کتابشناختی، قصد دارد با نمایش ساختارمند و معنایی موجودیت‌های داده‌ای بر اساس فرانمای تولید داده‌های ساختارمند و دیگر ابزارهای ساختارمندسازی داده‌های گوگل، دانش نهفته در بافت‌های داده‌ای گوناگون را بازنمون نماید. نگاره دانش، محصول یک پایگاه دانش است که توسط گوگل و خدمات آن مورد استفاده قرار می‌گیرد تا نتایج موتور جستجوی خود را با داده‌های گردآوری شده از منابع داده‌ای[8] گوناگون از بعد معناشناختی غنی سازد. نگاره دانش در ماه مه ۲۰۱۲ به موتور کاوش گوگل اضافه شد. در نگاره دانش هنگامی که یک موجودیت بازیابی می‌شود، موجودیت‌های مرتبط با آن نیز بازیابی، و روابط بین آن‌ها به‌صورت مصور، بازنمون می‌شود (Amit, 2012). البته این مهم زمانی تحقق می‌یابد که تمامی موجودیت‌ها و روابط میان آنها در یک بافت داده‌ای بازنمون شوند. بنابراین لازم است که از الگوهای هستی‌شناسانه هر بافت داده‌ای پیروی کند. در صورتی که تمام روابط بین موجودیت‌های بافت جهان کتابشناختی با استفاده از پیشینه‌های فراداده‌ای بازنمون شود، یک هستی‌شناسی کتابشناختی (فراداده‌ای) به‌وجود می‌آید. به عبارت دیگر، اگر پس از غنی‌سازی فرانمای تولید داده‌های ساختارمند از طریق ایجاد

1. Vocabularies
2. Microformats
3. Microdata
4. Resource Description Framework in Attributes (RDFa)
5. Javascript Object Notation for Linked Data (JSON-LD)
6.Things
7. Knowledge Graph
8. Data sources

واژگان‌های (فرانماهای) جدید برای دیگر انواع موجودیت‌های داده‌ای بافت میراث فرهنگی، همه موجودیت‌های زیرمجموعه آثار خلاقانه بر پایه واژگان‌های موجود توصیف شوند، دستیابی به یک هستی‌شناسی کتابشناختی (فراداده‌ای) تسهیل خواهد شد.

فرانمای تولید داده‌های ساختارمند برای هر موجودیت یک فرانما (واژگان) تدوین نموده است، اما پرسشی که مطرح می‌گردد آن است که انواع فرانماها تا چه میزان با انواع موجودیت‌های داده‌ای جهان کتابشناختی همخوانی دارند؟ به عبارت دیگر، کاربردپذیری فرانمای تولید داده‌های ساختارمند، متناسب با انواع موجودیت‌های داده‌ای بافت کتابشناختی به چه میزان است؟ بنابراین مسئله پژوهش این است که انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند، به چه میزان با انواع خاص منبع (اشیای محتوایی) بافت کتابشناختی همخوانی دارد و آیا تمامی انواع خاص را در برمی‌گیرد؟ برای پاسخ به این مسئله، هدف این پژوهش، تبیین میزان همخوانی انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با بازنمایی آن‌ها در فهرست اصطلاحات نوع خاص منبع استاندارد مارک ۲۱، به‌عنوان یکی از جامع‌ترین فهرست‌ها در این زمینه است.

مرور پیشینه‌های پژوهش

فرانمای تولید داده‌های ساختارمند دربرگیرنده واژگان ناهمگون بسیاری است که بافت‌های گوناگونی را تشکیل می‌دهد و این فرانما قصد دارد تقریباً همه انواع اشیای محتوایی وب را توصیف کند اما باید توجه داشت که بدون توجه به واژگان و اصلاحات تخصصی هر بافت توصیف تمامی اشیای محتوایی وب کامل میسر نخواهد شد، و بازنمون کامل دانش در هر بافت صورت نمی‌گیرد. بر این اساس، و با بررسی پژوهش‌های انجام شده در حوزه مورد پژوهش می‌توان پژوهش‌های پیشین را در دو گروه دسته‌بندی کرد. گروه اول پژوهش‌هایی هستند که تولید فرانماهای جدید متناسب با موجودیت‌های داده‌ای بافت داده‌ای هدف را مورد پژوهش قرار دادند، دیگر، پژوهش‌هایی که گسترش فرانماهای موجود برای توصیف جامع موجودیت‌های داده‌ای هر بافت داده‌ای را بررسی کردند. پژوهش‌های الدائج[1] (۲۰۱۵)؛ و محمدی‌استانی (۱۳۹۸) از جمله پژوهش‌هایی هستند که در گروه اول قرار دارند. گروه دوم شامل پژوهش‌های مویسل، بیزر

1. Aldaej

و پالهایم[1] (۲۰۱۵)؛ سیمسک، کارل، هولتسکنخت و فنسل[2] (۲۰۱۷)؛ جانگرا[3] (۲۰۱۸)؛ سینیور[4] (۲۰۱۸)؛ و آقاده (۱۳۹۷) می‌شود که به بررسی واژگان فرانمای تولید داده‌های ساختارمند برای توصیف موجودیت‌های داده‌ای بافت‌های مختلف پرداختند.

الدائچ (۲۰۱۵) پژوهشی با نام «محیط‌های یادگیری مجازی[5] پیشرفته مبتنی بر فرانمای تولید داده‌های ساختارمند و رسانه‌های اجتماعی»، با روشی تجربی و هدف ارائه رویکردی پیشرفته برای توصیف محتوای آموزشی با افزودن نشانه‌گذاری معنایی بر اساس اصطلاحات فرانمای تولید داده‌های ساختارمند و نیز یکپارچه‌سازی رسانه‌ها و شبکه‌های اجتماعی به منظور توسعه بافت یادگیری مجازی (وی‌ال‌ای) انجام داد. وی در این پژوهش اقدام به طراحی فرانمای اولیه وی‌ال‌ای و پیاده‌سازی آن با استفاده از نسخه ۷، سیستم مدیریت محتوای معنایی[6] (اس‌سی‌ام‌اس) دروپال[7] کرد، همچنین مفاهیم و اصطلاحات پیشنهادی پروژه خود از جمله نوع: دوره‌های آموزشی، نشست، ارزشیابی، آزمون و تکلیف[8] را برای توصیف محتوای وی‌ال‌ای‌ها با بهره‌جویی از مزایای فناوری‌های وب اجتماعی و معنایی، به دلیل پشتیبانی نکردن از توصیف منابع آموزشی در اصطلاحات فرانمای تولید داده‌های ساختارمند، ارائه کرد که در نهایت افزونه‌های یاد شده، در سلسله مراتب اصطلاحات این فرانما به‌عنوان زیر نوع، اکشن[9] قرار گرفتند. یافته‌ها همچنین حاکی از برتری مزیتی اس‌سی‌ام‌اس دروپال، نسبت به دیگر بسترهای موجود بود که امکان اشتراک محتوای آموزشی با شبکه‌های اجتماعی و ارائه ویژگی‌های پیشرفته رسانه‌ای، و پشتیبانی از فرانمای تولید داده‌های ساختارمند را برای ارتقای یافت‌پذیری شیء فراهم ساخت. همچنین «طراحی طرح‌واره قالب داده‌های خرد نسخه‌های خطی اسلامی-ایرانی و واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن» از محمدی‌استانی (۱۳۹۸) پژوهش دیگری از این گروه است. این پژوهش با استفاده از روش دلفی در طراحی فرانمای نسخه‌های خطی بافت اطلاعاتی ایرانی-اسلامی و روش تجربی در بررسی واکنش موتورهای کاوش وب انجام شد. ۲۱ صفت خاص نسخه‌های خطی بافت اطلاعاتی ایرانی_اسلامی در فرانمای تولید داده‌های ساختارمند و تعداد ۶ صفت برای افزودن به نوع «آثار خلاقانه» پیشنهاد گردید و واکنش

1. Meusel, Bizer & Paulheim
2. Şimşek, Kärle, Holzknecht & Fensel
3. Jangra
4. Senior
5. Virtual Learning Environments VLE
6. Semantic Content Management System (SCMS)
7. Drupal
8. Course, Session, Examination, test, Assignment
9. Action

موتورهای کاوش نسبت به پیشینه‌های مبتنی بر طرح‌واره پیشنهادی نسخه‌های خطی مثبت بود. پژوهش‌های انجام شـده فرانمایی دریک بافت هدف طراحی و به طراحان فرانمای تولید داده‌های ســاختارمند پیشــنهاد کرده‌اند که با به‌کارگیری آن در این فرانما بازنمون دانش به‌صورت کامل‌تر میسر می‌شود.

از پژوهش‌های گروه دوم که با هدف گســترش فرانماهای موجود انجام گرفته‌اند، مویســل و دیگران[1] (۲۰۱۵) پژوهشــی با عنوان «مطالعه‌ای در مقیاس وب درباره پذیرش و ســیر تکاملی اصــطلاحات فرانمای تولید داده‌های ساختارمند در طول زمان»، انجام دادند، که در آن با استفاده از خزشگرهای وب به بررســی روند پذیرش و ســیر تکاملی این اصــطلاحات پرداختند. آن‌ها برای انجام این پژوهش نســخه‌های مختلفی از این اصطلاحات را برای داده‌هایی که درصدها هزار صفحه وب در زمان‌های مختلف استقرار یافتند مقایســه کردند. این پژوهش بر روی داده‌های وب موجود در بنیاد کراول عمومی و به‌صــورت روش تجربی داده محور انجام گرفت. در این پژوهش مویســل و همکارانش از روش مشــاهده مســتقیم داده‌های منتشر شده توسط هزاران پذیرنده استاندارد فرانمای تولید داده‌های ساختارمند استفاده کردند. آن‌ها فرایند را به‌صــورت بالا به پایین و پایین به بالا اندازه‌گیری کردند. در مدل پذیرش بالا به پایین[2] این‌گونه تحلیل کردند که با چه ســرعت و تا چه اندازه‌ای تغییرات طرح توســط فراهم‌کنندگان داده پذیرفته شــده اســت. همچنین در ســیر تکاملی پایین به بالای[3] اصــطلاحات، بررســی کردند که اندازه تغییرات در فرانمای تولید داده‌های ساختارمند که توسط داده‌های توسعه‌یافته ایجاد شده، به چه میزان است. نتایج تحلیل تجربی آن‌ها نشــان داد که هر دو فرایند قابل مشــاهده اســت. پژوهش دیگری در این گروه «طراحی طرح‌واره داده‌های مســتند مبتنی بر روش داده‌های خرد و بررســی واکنش موتورهای کاوش وب به پیشــینه‌های مبتنی بر آن» بوده که توسط آقاده (۱۳۹۷) انجام شده است. در این پژوهش که به روش نیمه تجربی انجام شد، جامعه پژوهش شامل ۴۰۰ پیشینه فراداده‌ای مستند مبتنی بر روش داده‌های خرد و بستر نحوی جی‌سون_ال‌دی از کتابخانه دیجیتالی دانشــگاه علامه طباطبائی بود. یافته‌ها حاکی از آن بود که موتورهای کاوش قادر به نمایه‌سازی و بازیافت‌پذیرسازی تمامی پیشینه‌های جامعه پژوهش و ارزش‌های افزونه‌های مرتبط با داده‌های مستند طراحی شده بودند. سیمسک و دیگران (۲۰۱۷) در پژوهشی با عنوان «اعتبار معنایی حوزه خاصی از حاشیه‌نویسی فرانمای تولید داده‌های ساختارمند» با هدف اعتبار سنجی معنایی واژگان این فرانما و پاسخ به این مسـأله که فرانمای تولید داده‌های ســاختارمند قصــد دارد تقریباً همه چیز را در وب توصــیف کند و با

1. Meusel, & et al
2. Top-down adoption
3. Bottom-up evolution

توجه به اینکه وب بسیار بزرگ است، ناهمگونی در این فرانما ممکن است یک عارضه جانبی ایجاد کند، با استفاده از رویکرد ارزشیابی حاشیه‌نویسی فرانمای تولید داده‌های ساختارمند حوزه واژگان گردشگری را توصیف کردند.

از پژوهش‌های انجام شده دیگر در این گروه پژوهش جانگرا (۲۰۱۸) است. در مطالعه‌ای که روهیت جانگرا (۲۰۱۸) با عنوان «SchemaBibEx: یک ابتکار برای توصیف منابع کتابشناختی شبکه» انجام داده است بیان کرده که از قالب کتابشناختی مارک برای داده‌های کتابشناختی کتابخانه‌ها به خاطر تبدیل موجودیت‌های داده‌ای شبکه‌ای بسیار زیاد استفاده شده است. وی با این هدف که کتابخانه‌ها باید برای جذب کاربران از خدمات برنامه‌های شبکه‌ای به‌سمت سیستم کتابخانه‌ای راه‌حلی پیدا کنند، با استفاده از رویکرد گسترش توصیف منابع کتابشناختیِ فرانمای تولید داده‌های ساختارمند ابزاری قدرتمند تحت نام SchemaBibEx را به‌عنوان برای حل مشکل کتابخانه‌ها، مورد بحث قرار داده است. نتایج این مقاله با الگوی داده‌های پیوندی، اصطلاحات، ابزار و خدمات تجزیه‌وتحلیل شده است. جانگرا در پایان این پژوهش بیان کرده که SchemaBibEx جایگزینی برای مارک نیست، بلکه به یک پل ارتباطی دوطرفه بین داده‌های کتابشناختی کتابخانه‌ها و شبکه اطلاعاتی کاربران تبدیل خواهد شد. مطالعه دیگری که در این گروه قرار می‌گیرد پژوهشی با عنوان «همه را با یکدیگر ترکیب کنید: تطبیق منابع اصطلاحات پیایندها برای کشف داده‌های پیوندی» توسط آندرو سینیور (۲۰۱۸) است. وی این پژوهش را با هدف بررسی هم‌ترازی‌های احتمالی بین عناصری که در پیایندها استفاده می‌شود انجام داده است. او در این تحقیق با استفاده از اصطلاحات بیب‌فریم[1]، آردی‌ای رجیستری[2]، پرس‌اُاُ[3] و فرانمای تولید داده‌های ساختارمند، اصطلاحات مورد استفاده در پیایندها را با رویکردی تطبیقی مطالعه کرده است. یافته‌های این مطالعه حاکی از آن است که همبستگی‌هایی با درجات مختلف و به‌ندرت بین اصطلاحات چهار منبع وجود دارد اما ناهمگونی‌های مفهومی چالش اصلی این همترازی‌هاست.

با مطالعه پژوهش‌های انجام شده می‌توان گفت که پژوهش‌های گروه اول با طراحی و تولید فرانماهای جدید سعی در بهبود عملکرد فرانمای تولید داده‌های ساختارمند داشتند تا این فرانما در توصیف موجودیت‌های داده‌ای جدید در بافت‌های هدف بهتر عمل کند. در گروه دوم، این مطالعات به ارزیابی اصطلاحات فرانمای تولید داده‌های ساختارمند پرداختند و تعدادی از این پژوهش‌ها نیز گسترش واژگان

1. Bibframe
2. RDA Registery
3. Pressoo

فرانما را در موتورهای کاوش مورد مطالعه قرار دادند تا واکنش موتورهای کاوش را نسبت به آن‌ها دریابند. در پژوهش سینیور اشاره شده است که ناهمگونی مفهومی بین اصطلاحات چالش اصلی در ایجاد داده‌های پیوندی است. پژوهش جانگرا نیز گسترش توصیف منابع کتابشناختیِ فرانمای تولید داده‌های ساختارمند را مدنظر قرار داده است. بر این اساس پژوهش‌های گروه دوم به پژوهش حاضر نزدیک‌ترند اما با وجود انجام این مطالعات، خلأ پژوهشی‌ای که احساس می‌شود همگرایی بین داده‌های بافت میراث فرهنگی و فرانمای تولید داده‌های ساختارمند است. اینکه چگونه می‌توان کاربردپذیری فراداده‌های مبتنی بر فرانمای تولید داده‌های ساختارمند را بر اساس بافت کتابخانه‌ای و میراث فرهنگی بهبود بخشید. در این پژوهش سعی بر این است که با بررسی موجودیت‌های داده‌ای این فرانما بر اساس اصلاحات نوع خاص منبع مارک ۲۱ همگرایی بین آنها مشخص شود.

روش‌شناسی

پژوهش حاضر از نظر هدف یک پژوهش کاربردی است و به تبیین میزان همخوانی انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با فهرست اصطلاحات نوع خاص منبع استاندارد مارک ۲۱ می‌پردازد. در این پژوهش از روش تحلیل محتوا برای نیل به هدف آن استفاده شده است. بر اساس روش پژوهش، «واحدهای ثبت»، اصطلاحات نوع خاص منبع (ژانرهای) استاندارد مارک ۲۱ و موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند است و «واحدهای معنا»، استاندارد فراداده‌ای مارک ۲۱ و فرانمای تولید داده‌های ساختارمند هستند. جامعه پژوهش را انواع موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند تشکیل داده‌اند. ابزار گردآوری داده‌ها، یک سیاهه وارسی مبتنی بر عناصر فهرست اصطلاحات نوع خاص منبع مارک ۲۱ است که از آن برای بررسی میزان همخوانی با موجودیت‌های داده‌ای آثار خلاقانه در فرانمای تولید داده‌های ساختارمند در مرحله اول، و سپس با تمامی موجودیت‌های داده‌ای این فرانما، استفاده شد. دلیل انتخاب فهرست اصطلاحات نوع خاص منبع مارک این بوده است که این فهرست برای پوشش انواع منابع اطلاعاتی در بافت کتابشناختی از کامل‌ترین فهرست‌ها برای نوع خاص منبع این بافت است. این سیاهه با روش مشاهده طراحی شده تکمیل شد که برای این منظور هر یک از موجودیت‌های داده‌ای در فرانمای تولید داده‌های ساختارمند با اصطلاحات نوع خاص منبع مارک ۲۱ مقایسه و تطبیق داده شدند و با توجه به تعاریف، معانی و کارکردهای آنها، موجودیت‌های داده‌ای مرتبط شناسایی و استخراج شدند.

یافته‌های پژوهش

برای پاسخ به مسأله پژوهش، عناصر مطرح شده در فهرست اصطلاحات نوع خاص منبع مارک با انواع موجودیت‌های فرانمای تولید داده‌های ساختارمند تطبیق داده شدند تا مشخص شود هر یک از موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با کدام یک از اصطلاحات نوع خاص منبع مارک ۲۱ متناظر است. نخست، موجودیت‌های داده‌ای آثار خلاقانه و زیررده‌های آن، و سپس تمامی موجودیت‌های هسته و زیررده‌های آن‌ها در فرانمای تولید داده‌های ساختارمند با این اصطلاحات مقایسه شد، تا میزان همخوانی فهرست اصطلاحات نوع خاص منبع مارک با فرانمای تولید داده‌های ساختارمند مشخص شود. جدول ۱ همخوانی موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با عناصر فهرست اصطلاحات نوع خاص منبع مارک ۲۱ را نشان می‌دهد.

جدول ۱. همخوانی موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند با عناصر فهرست اصطلاحات نوع خاص منبع مارک

ردیف	اصطلاحات نوع خاص منبع مارک	معادل فارسی	اصطلاحات فرانمای تولید داده‌های ساختارمند	معادل فارسی
۱	abstract or summary	چکیده یا خلاصه	--	--
۲	art original	اصل اثر هنری	-	–
۳	art reproduction	باز تکثیر اثر هنری	-	–
۴	article	مقاله	Article	مقاله
۵	atlas	اطلس‌ها	Atlas	اطلس
۶	autobiography	خود سرگذشت‌نامه‌ها	-	–
۷	bibliographic data	داده‌های کتابشناختی	-	–
۸	bibliography	کتابشناسی	-	–
۹	biography	سرگذشت‌نامه	-	–
۱۰	book	کتاب	Book	کتاب
۱۱	calendar	تقویم	-	–
۱۲	catalog	فهرست	Datacatalog	فهرست داده‌ای
۱۳	chart	نمودار	-	–

۱۴	comedy	کمدی	Comedyevent (Event)	رویداد کمدی
۱۵	comic or graphic novel	رمان تصویری یا کمدی	Comicstory	داستان کمدی
۱۶	computer program	برنامه رایانه	-	–
۱۷	conference publication	انتشارات کنفرانس‌ها	-	–
۱۸	database	پایگاه داده	-	–
۱۹	dictionary	فرهنگ	-	–
۲۰	diorama	شهر فرنگ	-	–
۲۱	directory	راهنما	-	–
۲۲	discography	کتابشناسی مواد شنیداری	-	–
۲۳	document (computer)	سند (رایانه)	DigitalDocument	سند دیجیتالی
۲۴	drama	نمایشنامه	-	–
۲۵	encyclopedia	دایره المعارف	-	–
۲۶	essay	مقاله کوتاه ادبی	-	–
۲۷	festschrift	مجموعه آثار منتشر شده به افتخار یک محقق	PublicationEvent (event)	رویداد منتشر شده
۲۸	fiction	داستان	-	–
۲۹	filmography	کتابشناسی فیلم	-	–
۳۰	filmstrip	نوار فیلم	-	–
۳۱	finding aid	کمک‌های اولیه	-	–
۳۲	flash card	فلش کارت	-	–
۳۳	folktale		-	
۳۴	font	فونت	-	–
۳۵	game	بازی	Game	بازی
۳۶	globe	کره	-	––
۳۷	government publication	انتشارات دولتی	-	––
۳۸	graphic	گرافیک	-	––
۳۹	handbook	دستنامه	-	––
۴۰	history	تاریخ	-	––

۴۱	humor, satire	طنز، هجو	SatiricalArticle	مقاله طنز آمیز
۴۲	index	نمایه	-	--
۴۳	instruction	دستورالعمل	-	--
۴۴	interactive multimedia	چندرسانه‌ای تعاملی	-	--
۴۵	interview	مصاحبه	-	--
۴۶	issue	شمارگان	Publication issue	شمارگان انتشارات
۴۷	journal	نشریه علمی	-	
۴۸	kit	ترکیبی از اجزاء مختلف	Collection	مجموعه
۴۹	language instruction	آموزش زبان	Language (Intangible)	زبان
۵۰	law report or digest	خلاصه یا گزارش قانونی	-	-
۵۱	legal article	مقاله دولتی	-	--
۵۲	legal case and case notes	یادداشت‌های موردی و موارد حقوقی	LegalValueLevel (Intangible)	سطح اعتبار حقوقی
۵۳	legislation	وضع قانون	Legislation, legislationobject	وضع قانون
۵۴	letter	نامه	-	-
۵۵	loose-leaf	برگه‌های صحافی نشده	UpdateAction(Action)	به‌روزرسانی
۵۶	manuscript	نسخه خطی	-	-
۵۷	map	نقشه	Map	نقشه
۵۸	memoir	خاطرات	-	-
۵۹	microscope slide	اسلاید میکروسکوپ	-	-
۶۰	model	الگو	ProductModel(Product)	الگوی محصول
۶۱	motion picture	تصاویر متحرک	-	-
۶۲	newspaper	روزنامه	Newspaper	روزنامه
۶۳	nonmusical sound	صدای غیر موسیقایی	-	-
۶۴	novel	رمان	-	-
۶۵	numeric data	داده‌های شمارشی	Enumeration(Intangible)	شمارشی
۶۶	offprint	تیراژآپار (آف‌پرینت)	-	-
۶۷	online system or service	خدمات آنلاین	-	--

۶۸	patent	پروانه ثبت اختراع	-	--
۶۹	periodical	ادواری	Periodical	ادواری
۷۰	picture	عکس	ImageObject	شیء تصویری
۷۱	playing cards	کارت بازی	-	--
۷۲	poetry	شعر	-	--
۷۳	postcard	کارت پستال	-	--
۷۴	programmed text	کتاب خودآموز	-	--
۷۵	puzzle	پازل	-	--
۷۶	realia	واقع‌نما	-	--
۷۷	rehearsal	نمایش غیر موزیکال	-	--
۷۸	remote sensing image	تصویر سنجش از راه دور	ImageObject	شیء تصویری
۷۹	reporting	گزارش	Reaport	گزارش
۸۰	representational	اطلاعات تصویری	-	
۸۱	review	نقد	Review	نقد
۸۲	script	نوار	-	–
۸۳	series	فروست	CreativeWorkSeries	آثار خلاقانه فروستی
۸۴	short story	داستان کوتاه	-	–
۸۵	slide	اسلاید	-	–
۸۶	sound	صوت	Musicrecording	مواد موسیقایی
۸۷	speech	سخنرانی	-	–
۸۸	standard or specification	استاندارد یا	-	–
۸۹	statistics	آمار	-	–
۹۰	survey of literature	پژوهش در ادبیات	LiteraryEvent (Event)	رویداد ادبی
۹۱	technical drawing	نقاشی حرفه‌ای	DrawAction (Action)	عمل نقاشی
۹۲	technical report	گزارش فنی	TechArticle	مقاله فنی
۹۳	thesis	پایان‌نامه	Thesis	پایان‌نامه
۹۴	toy	اسباب‌بازی	ToyStore (organization & place)	اسباب‌بازی

				فروشی
۹۵	transparency	شفافیت	-	
۹۶	treaty	معاهده، پیمان	(Action) AgreeAction	توافقنامه
۹۷	videorecording	ضبط ویدئویی	VideoObject	شیء ویدئویی
۹۸	wall map	نقشه دیواری	MapCategoryType (Intangible)	انواع نقشه
۹۹	web site	وب‌سایت	Website	وب‌سایت
۱۰۰	yearbook	سالنامه	-	–

جدول فوق شامل ستون‌های اصطلاحات نوع خاص منبع مارک و معادل فارسی آنها و همچنین اصطلاحات فرانمای تولید داده‌های ساختارمند و معادل فارسی‌شان است. همان طور که در جدول نشان داده شده است، برای ۱۰۰ اصطلاح نوع خاص منبع مارک تنها ۳۶ معادل در فرانمای تولید داده‌های ساختارمند وجود داشت.

طبق داده‌های این جدول ۲۴ اصطلاح از اصطلاحات نوع خاص منبع مارک با آثار خلاقانه در فرانمای تولید داده‌های ساختارمند همخوانی کامل داشتند که شامل مقاله، اطلس، کتاب، فهرست، کتابشناسی مواد شنیداری، سند (رایانه)، بازی، طنز، شمارگان، مجموعه، وضع قانون، نقشه، گزارش فنی، پایان‌نامه، ضبط ویدئویی، نقشه دیواری و وب‌سایت می‌شوند. ۱۲ اصطلاح با سایر موجودیت‌های فرانما شامل رویداد، غیرملموس[1]، اقدام[2]، محصول، سازمان و مکان همخوانی تقریبی بودند[3] و برای ۶۴ اصطلاح هیچ معادلی یافت نشد.

برای اصطلاح abstract or summary در فرانمای تولید داده‌های ساختارمند تنها معادل abridged یافت شد که جزء «صفت‌های» موجودیت کتاب در این فرانماست و جزء نام موجودیت‌های داده‌ای نیست. همچنین برای sound در فرانمای تولید داده‌های ساختارمند اصطلاح audio وجود دارد اما جزء صفت‌هایی است که موجودیت‌ها را توصیف می‌کند و نزدیک‌ترین واژه برای sound موجودیت داده‌ای Musicrecording است. اصطلاح article در مارک شامل مقاله‌های مجلات، روزنامه‌ها و دائره‌المعارف‌ها است و در فرانمای تولید داده‌های ساختارمند دربرگیرنده مقاله‌های خبری و محتواهای تبلیغاتی است.

1. Intangible
2. Action
۳. نام موجودیت‌های داده‌ای غیر از آثار خلاقانه در جدول ۱ در مقابل هر موجودیت در پرانتز آورده شده است.

برنامه‌های رایانه‌ای در مارک با computer program مشخص شده ولی در فرانمای تولید داده‌های ساختارمند با computer languageکه به زبان‌های برنامه‌نویسی کامپیوتر اشاره دارد آمده است و با هم متفاوتند. اصطلاح discography در نوع خاص منبع مارک به کتابشناسی مواد شنیداری اشاره کرده است و Musicrecording نزدیک‌ترین واژه به آن در فرانمای تولید داده‌های ساختارمند است که ارتباط بسیار دوری است و نمی‌توان آنان را همخوان شمرد.

در مورد toy در فرانما به محل آن توجه شده که می‌تواند خود toy را نیز در برگیرد. همچنین برای اصطلاح technical drawing در مارک، عمل نقاشی DrawAction در نظر گرفته شده است که نزدیک‌ترین معادل برای آن است. بر این اساس و طبق یافته‌های این پژوهش ۳۶ درصد اصطلاحات نوع خاص منبع مارک با فرانمای تولید داده‌های ساختارمند همخوان است.

نتیجه

الگوها و استانداردهای کتابخانه‌ای به‌واسطه پیشینه غنی و تجربه گسترده در استفاده از فراداده‌ها، به بلوغ نسبی در توصیف موجودیت‌های داده‌ای در محیط شبکه‌ای و غیرشبکه‌ای رسیده‌اند، و از آنجایی که طراحان موتورهای کاوش برای ایجاد داده‌های ساختارمند به منظور پردازش داده‌ها به‌صورت معنایی کوشش‌هایی را برای توصیف موجودیت‌های داده‌ای وب صورت داده‌اند، بنابراین برای اینکه داده‌های بافت کتابشناختی توسط موتورهای کاوش به بهترین شکل و به‌صورت معنایی پردازش، جستجو و بازیابی شوند ضروری است که بین انواع موجودیت‌های داده‌ای که موتورهای کاوش معرفی کرده‌اند و نیز انواع داده‌های کتابشناختی همخوانی وجود داشته باشد. از این‌روی فرانمای تولید داده‌های ساختارمند که از سوی موتورهای کاوش معرفی شده است می‌بایست بتواند داده‌های بافت میراث فرهنگی و در اینجا بافت کتابشناختی را به‌صورت کامل توصیف نماید. بر این اساس در این پژوهش فهرست اصطلاحات نوع خاص منبع مارک (مارک به‌عنوان استاندارد فراداده‌ای) با موجودیت‌های فرانمای تولید داده‌های ساختارمند تطبیق داده شد تا بتوان میزان همخوانی موجودیت‌های این فرانما را بر اساس نوع خاص منبع‌های موجود در جهان کتابشناختی تعیین کرد. بر اساس یافته‌های این پژوهش که در جدول ۱ ارائه شد، ۲۴ اصطلاح از نوع خاص منبع مارک با آثار خلاقانه در فرانمای تولید داده‌های ساختارمند همخوانی داشت. ۱۲ اصطلاح با سایر موجودیت‌های فرانما شامل رویداد، غیرملموس، اقدام، محصول، سازمان و مکان همخوان بود و برای ۶۴ اصطلاح هیچ معادلی یافت نشد و در واقع تنها ۳۶ درصد از موجودیت‌های فرانمای تولید داده‌های ساختارمند با اصطلاحات نوع خاص منبع مارک همخوانی داشت.

فرانمای تولید داده‌های ساختارمند یک استاندارد فراداده‌ای محسوب می‌شود و در استفاده از صفت‌ها دقیق و گسترده عمل کرده و سعی کرده است متناسب با موجودیت‌های داده‌ای وب باشد و مارک نیز غنی‌ترین استاندارد فراداده‌ای بافت کتابشناختی است که فهرستی از غنی‌ترین نوع خاص منبع را پوشش می‌دهد. چنانچه فرانمای تولید داده‌های ساختارمند با فهرست اصطلاحات نوع خاص منبع مارک تطابق نداشته باشد، نمی‌تواند موجودیت‌های داده‌ای بافت کتابشناختی را به‌طور جامع پوشش داده، و بنابراین بازنمون دانش محقق نخواهد شد. بدین ترتیب، کاربردپذیری این فرانما در بافت کتابشناختی کاهش خواهد یافت.

در مقایسه یافته‌های پژوهش‌های پیشین با این پژوهش می‌توان در پژوهش‌های خارجی و در گروه دوم به پژوهش سینیور (۲۰۱۸) اشاره کرد که به تطبیق منابع اصطلاحات پایندها برای کشف داده‌های پیوندی پرداخته و اشاره کرده است که ناهمگونی مفهومی بین اصطلاحات چالش اصلی در ایجاد داده‌های پیوندی است. او در این پژوهش، با استفاده از اصطلاحات بیب‌فریم، آردی‌ای رجیستری، پرس‌اُاُ و فرانمای تولید داده‌های ساختارمند، اصطلاحات مورد استفاده در پایندها را با رویکردی تطبیقی مطالعه کرده است. همچنین پژوهش جانگرا (۲۰۱۸) با رویکرد گسترش توصیف منابع کتابشناختی در فرانمای تولید داده‌های ساختارمند انجام شده است. در پژوهش حاضر هم از فهرست اصطلاحات نوع خاص منبع مارک برای تطبیق با موجودیت‌های فرانمای تولید داده‌های ساختارمند استفاده شد. در این پژوهش مشخص شد فرانمای تولید داده‌های ساختارمند چنانچه اصطلاحات مارک را بکار گیرد در بازنمون بافت کتابخانه‌ای موفق‌تر عمل خواهد کرد. این پژوهش قصد داشت با همخوانی استانداردهای بافت میراث فرهنگی (با تأکید بر بافت کتابخانه‌ای) با فرانمای تولید داده‌های ساختارمند توصیف منابع کتابشناختی را در این فرانما گسترش دهد. بر مبنای یافته‌های این پژوهش توصیف موجودیت‌های داده‌ای بر اساس فرانمای تولید داده‌های ساختارمند برای بهره‌گیری از استانداردهای فراداده‌ای و کارکردپذیری مدنظر بافت میراث فرهنگی به‌ویژه بافت کتابشناختی به نظر کافی نیست.

بر پایه یافته‌های پژوهش حاضر، برای ۶۴ اصطلاح فهرست اصطلاحات نوع خاص منبع مارک در فرانمای تولید داده‌های ساختارمند هیچ معادلی یافت نشد. با نگاهی اجمالی و به‌صورت ظاهری، با توجه به اینکه این فرانما بر روی اشیای محتوایی وب متمرکز است، می‌توان گفت به تعریف برخی از موجودیت‌های داده‌ای بافت کتابشناختی در فرانمای تولید داده‌های ساختارمند نیازی نبوده و ضروری نیست اما برای توصیف و سازماندهی اشیای دیجیتالی با توجه به اینکه بیشتر موجودیت‌های داده‌ای موجود در وب نسخه دیجیتالی شده اشیای محتوایی فیزیکی هستند، بنابراین ویژگی‌های هر دو باید مورد توجه قرار گیرد و نیز به

دلیل اینکه فرانمای تولید داده‌های ساختارمند بخشی از موجودیت‌های داده‌ای بافت کتابشناختی را در نظر نگرفته است اگر فرانمای تولید داده‌های ساختارمند بخواهد توصیف اشیاء وب را تحت پوشش قرار دهد و بازنمون دانش را تکمیل کند باید برای توصیف تمام نوع خاص منبع‌های بافت کتابشناختی فرانما (اصطلاحات) داشته باشد در غیر این صورت تحقق بازنمون دانش به‌صورت کامل میسر نخواهد شد. این پژوهش با تأکید بر بافت کتابخانه‌ای صورت گرفت. با توجه به اینکه گسترش توصیف موجودیت‌های داده‌ای وب در هر بافت برای بازنمون دانش مورد نیاز است، پیشنهاد می‌شود پژوهش‌هایی با تأکید بر بافت آرشیو و بافت موزه از زیرمجموعه‌های بافت میراث فرهنگی صورت پذیرد، و انواع موجودیت‌های مربوط به این دو بافت با موجودیت‌های داده‌ای فرانمای تولید داده‌های ساختارمند تطبیق داده شود تا خلأهای موجود شناسایی شده و زمینه برای بازنمون کامل دانش در بافت میراث فرهنگی فراهم شود.

منابع

آقاده، سمیرا (۱۳۹۷). *طراحی طرح‌واره داده‌های مستند مبتنی بر روش داده‌های خرد و بررسی واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن*. پایان‌نامه کارشناسی ارشد. دانشکده روانشناسی و علوم تربیتی. دانشگاه علامه طباطبائی.

محمدی استانی، مرتضی (۱۳۹۸). *طراحی طرح‌واره قالب داده‌های خرد نسخه‌های خطی اسلامی-ایرانی و واکنش موتورهای کاوش وب به پیشینه‌های مبتنی بر آن*. رساله دکتری. دانشکده روانشناسی و علوم تربیتی. دانشگاه اصفهان.

Aldaej, A. A. (2015). *An enhanced semantic VLE based on schema. org and social media* (Doctoral dissertation, University of Surrey). From: https://epubs.surrey.ac.uk/807070/1/Aldaej%20PhD%20Thesis.pdf (accessed 28 March 2021)

Amit, S. 2012. *Introducing the knowledge graph: Things, not strings. Official Blog (of Google).* https://googleblog.blogspot.se/2012/05/introducing-knowledge-graph-things-not.html.(Accessed 15 May. 2018)

Google developers .2017. *Introduction to structured data.* https://developers.google.com/search/docs/guides/intro-structured-data/.(Accessed 15 July. 2018)

Jangra, R. 2018. *SchemaBibEx: An Initiative for Networked Bibliographic Resource Description*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3149541. (accessed 5 June. 2018)

Meusel, R., Bizer, C., & Paulheim, H. 2015. *A web-scale study of the adoption and evolution of the schema. org vocabulary over time*. In *Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics* (p. 15). ACM. From: https://doi.org/10.1145/2797115.2797124

Schema.org (2019). *Getting started with schema.org using Microdata*. Retrieved July 19, 2019

from: http://schema.org/docs/gs.html

Senior, A. (2018). Bringing It All Together: Mapping Continuing Resources Vocabularies for Linked Data Discovery. *The Serials Librarian*, 74(1-4), 94-101, from: https://doi.org/10.1080/0361526X.2018.1428463

Şimşek, U., Kärle, E., Holzknecht, O., & Fensel, D. (2017, June). *Domain specific semantic validation of schema. org annotations*. In *International Andrei Ershov Memorial Conference on Perspectives of System Informatics* (pp. 417-429). Springer, Cham, from: https://doi.org/10.1007/978-3-319-74313-4_31

Vizine-Goetz, D., Childress, E., & Houghton, A. (2005, September). Web services for genre vocabularies. In *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC 2005), Madrid, Spain (September 12-15 2005)*. https://dcpapers.dublincore.org/pubs/article/view/823/819

A Study on the Corresponding between Schema.org's Item Types with MARC Genre Terms List

Mahsa Fardehosseini
Islamic Azad University, Tehran
Seyed Mahdi Taheri*
Allameh Tabataba'i University
Nadjla Hariri
Fahimeh Babalhavaeji
Fatemeh Nooshinfard
Islamic Azad University, Tehran

Abstract

Introduction: The purpose of this study was to determine the degree of correspondence between data types of Schema.org and the representation of them in the MARC21 Genre Terms List.

Methodology: This research was an applied research based on a content analysis and the research population was a variety of data entities of the Schema.org. To collect data, a checklist-based tool on MARC Genre Terms List was used to check the degree of correspondence in Schema.org, and the structured observation method was used to initially identify each of the CreativeWork and then other entities in the Scheme.org were analyzed and the types of related data entities were identified

Findings: The findings of this study showed that there are 64 source-specific terms in the MARC that have no equivalent data entity in the Schema.org, and only 36% of the MARC Genre Terms List are consistent with the Schema.org. This means that for the 100 MARC Genre Terms List, there were only 36 equivalents in the Schema.org. Twenty-four terms of the MARC Genre Terms List were consistent accurately with the CreativeWork in Schema.org, and 12 were consistent with its other data entities.

Conclusion: To describe and organize digital objects, both properties must be taken into account because most data entities in the web are digitized versions of physical content objects and also considering that Schema.org does not consider some of the data entities in the bibliographic context, it must design the schema to describe all terms of a particular type of cultural heritage context (including the bibliographic context) for removing knowledge gap. It is recommended that specialized refinements and vocabularies of other sections of the cultural heritage (museum and archive context) be adapted to the types of the Schema.org to provide a thorough representation of knowledge in the context of cultural heritage.

Keywords: Bibliographic context, CreativeWorks, Cultural heritage context, MARC Genre Terms List, Schema.org

* . taherismster@gmail.com